2020 Second International Conference on Transdisciplinary AI (TransAI)

Essential Question: 'Equal or Equivalent Entities?' About Two Things as Same, Similar, or Different

1st Anousha Athreya
Brain Health Alliance
Ladera Ranch, CA, USA
aathreya@bhavi.us

2nd S. Koby Taswell
Brain Health Alliance
Ladera Ranch, CA, USA
ktaswell@bhavi.us

3rd Sohyb Mashkoor
Brain Health Alliance
Ladera Ranch, CA, USA
smashkor@bhavi.us

4th Carl Taswell
Brain Health Alliance
Ladera Ranch, CA, USA
ctaswell@bhavi.us

***Abstract*—We discuss definitions of entities, equality, and equivalence as used by a transdisciplinary diversity of research fields including mathematics, statistics, computational linguistics, computer programming, knowledge engineering, and music theory. Declaring definitions for these concepts in the situational context of each domain specific field supports the essential question 'Equal or equivalent entities?' about two things as same, similar, related, or different for that field. Pattern recognition performed by artificial intelligence applications can be described as the automated process of answering this fundamental question about the similarity or difference between two things.**

***Index Terms*—DREAM principles, equivalent entities, semantic web, knowledge engineering, ontology, plagiarism.**

I. Contrasts and Comparisons

As a species, humans have evolved to understand the world around us by adapting for survival and learning to differentiate friend from foe. But how do we discern when two things are same, similar, related, or different? Contrasts distinguish one thing from another with emphasis on the differences, while comparisons examine analogies, similarities, or relative rankings of related or similar things. General concepts about contrasts and comparisons can be applied to analytic methods for diverse scientific research fields. In biology, comparisons between two or more creatures enable their classifications into the different species, genus, family, *etc.*, of the taxonomic hierarchy. In mathematics, comparisons between two or more numbers, or two or more functions over a domain with values in a range, enable their analysis with the formal definitions for basic operations such as $\min(\cdot), <, \leq, =, \geq, >, \max(\cdot)$, *etc.* In statistics, formal methods exist with hypothesis testing for the null versus alternative hypotheses when calculating the differences between two sets of data. In knowledge engineering, words listed in different vocabularies, thesauri, and ontologies representing concepts and ideas can be mapped to each other as same, similar, related, or different, and then applied to the review of published literature. This analysis remains an important requirement for integrity and ethics in scholarly research to evaluate novelty versus plagiarism [1].

II. Equal or Equivalent Entities?

From the Merriam-Webster dictionary definition of entity, an entity is "the existence of a thing as contrasted with its attributes". In the context of the PORTAL-DOORS Project for the semantic web, an entity has been defined as "the object of interest considered by the registrant to be the resource whether concrete or abstract, online or offline, semantic or lexical, real or virtual" [2]. In this context, two entities are considered equivalent when there exists sufficient semantic similarity despite lexical differences including differences in natural language (*i.e.*, English, French, Russian, *etc.*). The term 'equivalent entities' and the question 'Equal or equivalent entities?' both represent the essential enquiry of identifying and characterizing two entities as the same, similar, related, or different from each other. This equivalent entities principle remains at the center of the DREAM principles with summarizing phrase "Discoverable Data with Reproducible Results for Equivalent Entities with Accessible Attributes and Manageable Metadata". Quoting from Dutta *et al.* [3], *"We emphasize that science will be neither reproducible nor fair without recognition, acknowledgment, attribution and citation of equivalent entities regardless of whether those equivalent entities are considered to be scientific hypotheses, scientific experiments, scientific data, scientific results or published articles in the scientific literature."*

We examine this question about equivalent entities in a variety of fields including mathematics, statistics, computational linguistics, computer programming, and ontology engineering. In mathematics, logical and numerical entities are defined formally within a context of axioms, postulates, and theorems, which provide the basis for more complicated relations, calculations, and analyses. Euclid, the father of geometry, defined five central axioms including concepts related to equality and equivalence [4]. Various definitions for equality and equivalence exist throughout mathematics and their applications from basic arithmetic to sophisticated computational algorithms [5]. In programming languages such as C# [6], the equality operator '==' for reference types compares the reference identities and indicates that both the type and value are the same, whereas the '.Equals()' method is a virtual method that can be overriden and thus mapped to define an equivalence operator for the type and value of the class-defined object. In biomedical statistics, equivalence can be tested with a 'margin' denoted by δ, where two treatments are tested for similar effects within a range of values specified by the margin δ [7]. In computational linguistics, an equivalence studied for natural language processing examines whether two languages use the same set of abstract syntax [8].

978-1-7281-8699-3/20/$31.00 ©2020 IEEE
DOI 10.1109/TransAI49837.2020.00028

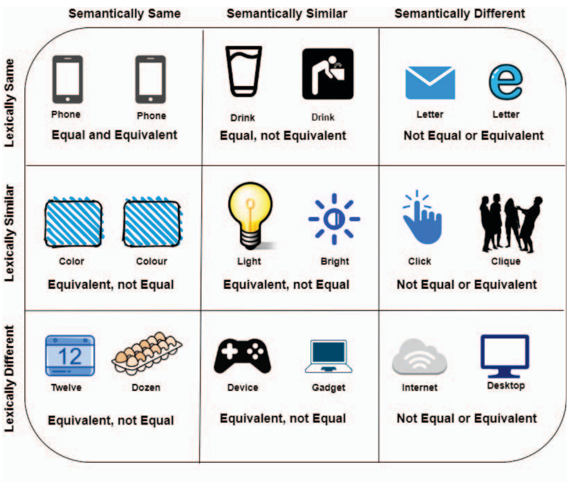

Figure 1. Evaluation of Equal and Equivalent Instances with Same, Similar, and Different

In the examples above as well as those from other diverse fields of science, when the question 'Equal or equivalent entities?' has been asked and answered with uncertainty as an equivalence, then the corollary question 'True or false equivalence?' must also be considered. If either the data about the two entities are incorrect or the analysis algorithm is flawed and invalid, then the result of equivalence must be considered a false equivalence rather than a true equivalence [9]. When reviewing results from automated artificial intelligence applications, human experts should apply both logical reason and intuitive common sense when deciding between a true equivalence or a false equivalence. Figure 1 presents pairs of pictographs in the cells of a $3x3$ table with 9 cases demonstrating simple examples for contrast and comparison of lexically and semantically same, similar, and different.

III. Same, Similar, Related, or Different?

Defining same, similar, related, and different in order to determine equality or equivalence of entities must be declared in the context of a problem-oriented domain for a specific field of scientific research. This context imbues meaning and sense for the hierarchies and associations found within domain-specific ontologies not only for entities but also for the relationships between entities. Then equivalence between concepts can be established for both entities and relationships [10]. In music theory as a dramatic example of the importance of situational context, a pitch is heard in relation to the other pitches around it. Even if using the same set of notes, by changing the starting position in the sequence of notes, the perceived mode and sound changes. Thus, musical scales can be similar and related, but not the same, despite having the same set of notes [11]. In general, comparisons between two entities for equality or equivalence can be made by observing for the presence of common shared attributes between the two entities using similarity measures to characterize identifiable attributes [12]. The greater the number of similar attributes the two entities share in common, the greater the overall similarity between the two entities. Given a declared minimum threshold for the number of shared attributes required for similarity, the two entities can be considered similar or different, respectively, when the number observed is above or below the threshold. If all of the observed attributes are similar for the two entities, then certainly they are equivalent. But whether they are also equal depends on the context of the domain specific question.

References

[1] S. K. Taswell, C. Triggle, J. Vayo, *et al.*, "The hitchhiker's guide to scholarly research integrity," in *2020 ASIS&T 83rd Annual Meeting*, (Oct. 22, 2020), Wiley, 2020.

[2] C. Taswell, "A distributed infrastructure for metadata about metadata: The HDMM architectural style and PORTAL-DOORS system," *Future Internet*, vol. 2, no. 2, pp. 156–189, 2010, In Special Issue on Metadata and Markup., ISSN: 1999-5903. DOI: 10.3390/FI2020156.

[3] S. Dutta, P. Kowshik, A. Ambati, *et al.*, "Managing scientific literature with software from the PORTAL-DOORS Project," in *2019 IEEE 15th International Conference on eScience (eScience)*, (Sep. 24, 2019), San Diego, California: IEEE, Sep. 2019. DOI: 10.1109/eScience.2019.00081.

[4] I. Mueller, "Euclid's elements and the axiomatic method," *The British Journal for the Philosophy of Science*, vol. 20, no. 4, pp. 289–309, 1969.

[5] K. Raczkowski and P. Sadowski, "Equivalence relations and classes of abstraction," *Formalized Mathematics*, vol. 1, no. 3, pp. 441–444, 1990.

[6] J. Albahari and B. Albahari, *C# 7.0 in a nutshell: The definitive reference*. " O'Reilly Media, Inc.", 2017.

[7] E. Walker and A. S. Nowacki, "Understanding equivalence and noninferiority testing," *Journal of general internal medicine*, vol. 26, no. 2, pp. 192–196, 2011.

[8] R. Hausser, *Foundations of Computational Linguistics*. Springer, 2001.

[9] S. Choksi, P. Hong, S. Mashkoor, *et al.*, "NPDSLINKS: Nexus-PORTAL-DOORS-Scribe Learning Intelligence aNd Knowledge System," in *2020 IEEE 2nd International Conference on Transdisciplinary Artificial Intelligence (TransAI)*, (Sep. 21, 2020).

[10] S. Zhang and O. Bodenreider, "Comparing associative relationships among equivalent concepts across ontologies," *Studies in health technology and informatics*, vol. 107, no. 0 1, p. 459, 2004.

[11] A. Forte, "A theory of set-complexes for music," *Journal of Music Theory*, vol. 8, no. 2, p. 136, 1964. DOI: 10.2307/843079.

[12] M. Holub, O. Proksa, and M. Bieliková, "Detecting identical entities in the semantic web data," in *International Conference on Current Trends in Theory and Practice of Informatics*, Springer, 2015, pp. 519–530.